

# 扩增子实验分析及试剂耗材

2019 年 11 月

## 目录

<b>1. 实验流程</b>	<b>3</b>
1.1 DNA 提取	3
1.2 PCR 扩增及产物电泳检测	3
1.3 Pooling 及切胶纯化	4
1.4 建库及测序	4
<b>2. 分析流程</b>	<b>5</b>
2.1 测序数据处理	5
2.2 OTU 聚类及物种注释	5
2.3 OTU 统计	6
2.4 物种群落分析	6
2.5 Alpha 多样性分析	7
2.6 Beta 多样性分析	7
2.7 物种差异分析	8
2.8 环境因子关联分析	8
2.9 网络与预测分析	9
2.10 16S 功能预测分析	10
<b>3. 试剂、耗材及仪器</b>	<b>11</b>
3.1 试剂列表	11
3.2 常用耗材	11
3.3 常用仪器	12

# 1. 实验流程

## 1.1 DNA 提取

使用各类样本对应的 DNA 提取试剂盒/CTAB 法/SDS 法（详见 DNA 质检报告）进行基因组 DNA 抽提后，利用 Thermo NanoDrop One 检测 DNA 的完整性、纯度和浓度。

## 1.2 PCR 扩增及产物电泳检测

以基因组 DNA 为模板，根据测序区域的选择，使用带 barcode 的特异引物及 TaKaRa Premix Taq® Version 2.0(TaKaRa Biotechnology Co., Dalian, China)进行 PCR 扩增。

### 1.2.1 引物对应区域

16S V4 区引物（515F 和 806R）：鉴定细菌多样性；

18S V4 区引物（528F 和 706R）：鉴定真核微生物多样性；

ITS1 区引物（ITS5-1737F 和 ITS2-2043R）：鉴定真菌多样性；

此外，扩增区域还包括：16S V3-V4/16S V4-V5；古菌 16S V4-V5；18S V5 和 ITS2 区；功能基因对应引物等。

### 1.2.2 PCR 反应体系

试剂名称	用量
2x Premix Taq	25 $\mu$ l
Primer-F（10 $\mu$ M）	1 $\mu$ l
Primer-R（10 $\mu$ M）	1 $\mu$ l
DNA	50 ng
Nuclease-free water	Add to 50 $\mu$ l

### 1.2.3 PCR 反应条件

1) 94°C 5min

2) 30 cycle of

94°C 30s

52°C 30s

72°C 30s

3) 72°C 10min

4) 4°C Hold

每个样本进行 3 个重复，并将同一样本的 PCR 产物进行混合，PCR 仪：BioRad S1000 (Bio-Rad Laboratory, CA)。

#### 1.2.4 PCR 产物电泳检测

用 1%琼脂糖凝胶电泳检测 PCR 产物的片段长度和浓度，主带长度在正常范围内（例如 16S V4: 290-310bp/16S V4-V5: 400-450bp 等）的样品可用于进一步的实验。

### 1.3 Pooling 及切胶纯化

利用 GeneTools Analysis Software (Version4.03.05.0, SynGene)对 PCR 产物进行浓度对比后，按照等质量原则计算各样品所需体积，将各 PCR 产物进行混合。使用 E.Z.N.A.® Gel Extraction Kit (Omega, USA)凝胶回收试剂盒回收 PCR 混合产物，TE 缓冲液洗脱回收目标 DNA 片段。

### 1.4 建库及测序

#### 1.4.1 建库

按照 NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® （New England Biolabs, USA）标准流程进行建库操作。

#### 1.4.2 测序

使用 **Illumina Nova 6000** 平台对构建的扩增子文库进行 PE250 测序。(Guangdong Magigene Biotechnology Co., Ltd. Guangzhou, China)。

## 2. 分析流程

### 2.1 测序数据处理

(1) Paired-end Raw Reads 数据过滤：利用 fastp (an ultra-fast all-in-one FASTQ preprocessor, version 0.14.1, <https://github.com/OpenGene/fastp>) 分别对双端的 Raw Reads 数据进行滑窗质量剪裁(-W 4 -M 20), 同时, 根据序列首尾两端的引物信息, 利用 cutadapt 软件(<https://github.com/marcelm/cutadapt/>) 去除引物, 得到质控后的 paired-end Clean Reads。

(2) Paired-end Clean Reads 拼接：对于双端测序数据, 需要根据 PE reads 之间的 overlap 的关系, 利用 usearch -fastq\_mergepairs (V10, <http://www.drive5.com/usearch/> 预设参数包含最小 overlap 长度设置为 16bp, 拼接序列的 overlap 区允许的最大错配 5bp 等), 过滤不符合的 Tags, 获得原始的拼接序列(Raw Tags)。

(3) Raw Tags 序列质量过滤：利用 fastp (an ultra-fast all-in-one FASTQ preprocessor, version 0.14.1, <https://github.com/OpenGene/fastp>) 对 Raw Tags 数据进行滑窗质量剪裁(-W 4 -M 20), 得到有效的拼接片段(Clean Tags)。

### 2.2 OTU 聚类及物种注释

(1) OTU 聚类：OTU, 即 operational taxonomic units, 是微生物学研究中最常见的术语之一。平台提供以下三种方法, 默认**聚类方法为 UPARSE**:

1. **UPARSE** (RC Edgar. highly accurate OTU sequences from microbial amplicon reads. Nature methods, 2019, 10(10): 996);
2. **UNOISE3** (RC Edgar. UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon read. bioRxiv, 2016);
3. **UCLUST** (RC Edgar. Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 2010, 26(19):2460-2461)。

(2) 代表性序列物种注释：利用 usearch -sintax 将每个 OTU 的代表序列与 SILVA(16S)、RDP(16S)、Greengenes(16S)、SILVA (18S)、Unite(ITS)数据库进行

比对获得物种注释信息（设定置信度阈值默认为 0.8，默认数据库 SILVA(16S)、SILVA (18S)和 Unite(ITS)），从而达到了解所有序列物种来源的目的。物种注释的 taxonomy 结果分为 7 个层级，分别为界（kingdom, L1）、门（phylum, L2）、纲（class, L3）、目（order, L4）、科（family, L5）、属（genus, L6）和种(species, L7)。

(3) 污染 OTU 去除：去除注释为叶绿体或线粒体（16S 扩增子）以及不能注释到界级别的 OTU 及其 Tags，得到各样品最终用于分析的有效 Tags 序列数(No. of seqs)及 OTU 分类学综合信息表(OTU\_table)。

## 2.3 OTU 统计

(1) OTU\_Table: 基于上述去 singleton OTU、嵌合体及污染 OTU 后的 OTU\_table，统计样本或分组所含 reads 数及 OTU 数。

(2) Pan\_Core 物种分析: 使用 R 软件(V5.1.3)统计不同样品数中目标分类水平的并集数(Pan)及交集数(Core)，来评估样本量是否充足。

## 2.4 物种群落分析

(1) 物种群落结构: 使用 R 软件进行共有及特有物种统计、群落组成分析及物种丰度聚类分析。

(2) 系统进化分析:

1. 单样本-各分类水平系统进化分析: a) 基于样本内各 OTU 的物种系统进化关系及相对丰度信息，使用 KRONA 软件(<http://sourceforge.net/projects/krona/>)对单个样本的物种注释结果进行可视化；b) 为了快速直观的挖掘研究样品中的物种组成及丰度信息，我们使用 GraPhlAn 软件(<http://huttenhower.sph.harvard.edu/graphlan>)得到基于 GraPhlAn 的单个样本 OTU 注释圈图。
2. 各分类水平-所有样本系统进化分析(默认参数): 选取总体相对丰度排在前 20 的 OTU 代表性序列，使用 FastTree 软件无限比对建树，同时结合

每个 OTU 的相对丰度及其代表序列的物种注释置信度信息,使用 ggtree 软件包进行可视化展示。

(3) 组间群落结构差异显著性分析:使用 R 软件 vegan 及 pegas 包的 anosim 函数、mrpp 函数、adonis 函数和 Amova 进行 Anosim、MRPP、Adonis 及 Amova 显著性分析。

## 2.5 Alpha 多样性分析

(1) Alpha 多样性指数统计(默认参数):基于 OTU 丰度表,使用 usearch -alpha\_div (V10, <http://www.drive5.com/usearch/>) 进行 14 种多样性指数(richness、chao1、shannon\_2、shannon\_2、shannon\_e、shannon\_10、jost、jost1、simpson、dominance、equitability、robbins、berger\_parker、reads、buzas\_gibson)的计算。

(2) 稀释曲线(默认参数):基于 OTU 丰度表,使用 usearch -alpha\_div\_rare (V10, <http://www.drive5.com/usearch/>) 进行上述 14 种多样性指数的稀释曲线计算。

(3) Rank Abundance 曲线(默认参数):基于 OTU 丰度表,使用自建 python 脚本 (python v2.7), 进行 Rank Abundance 计算。

(4) 指数组间差异检验:使用 R 软件进行 alpha 多样性指数组间差异分析, alpha 多样性指数组间差异分析会分别进行有参数检验和非参数检验,如果只有两组,选用 Student's t-test 或 wilcox 秩和检验,如果多于两组,选用的是 Kruskal-Wallis 秩和检验或 one-way ANOVA。

## 2.6 Beta 多样性分析

(1) NMDS 分析(默认参数):基于 OTU 丰度表,使用 R 软件的 vegan 软件包配合 9 种距离算法(bray\_curtis、Euclidean、abund\_jaccard、Canberra、chisq、chord、gower、weighted\_unifrac、unweighted\_unifrac)进行分析。

(2) PCA 分析(默认参数):基于 OTU 丰度表,使用 R 软件的 prcomp 函数进行分析。

(3) PCoA 分析(默认参数): 基于 OTU 丰度表, 使用 R 软件的 `vegan` 软件包配合上述 9 种距离算法进行分析。

(4) 聚类分析(默认参数): 基于 OTU 丰度表, 基于 9 种 beta 多样性距离矩阵, 采用 UPGMA(Unweighted Pair-group Method with Arithmetic Mean)聚类分析方法构建样品的聚类树。

(5) 样本距离 Heatmap 分析(默认参数): 使用 R 软件的 `vegan` 软件包合并上述 9 种距离算法及 `hclust` 函数进行热图聚类分析。

## 2.7 物种差异分析

(1) LEfSe 分析: 基于均一化的各个物种等级的丰度表 `OTU_table`, 使用 LEfse 软件进行组间物种的差异显著分析, 首先使用 non-parametric factorial Kruskal-Wallis(KW)sum-rank test 检测不同分组间丰度差异显著的物种, 然后用成组的 Wilcoxon 秩和检验来进行两两组间差异性判断, 最后用线性判别分析(LDA)来实现降维和评估差异显著物种的影响大小(即为 LDA Score), 默认设置 LDA Score 的筛选值为 2, 结果即为各组内的 biomarker。

(2) 组间差异显著性检验—多组比较: 基于 OTU 丰度表, 使用 R 软件进行 Kruskal-Wallis 秩和检验或 One-way ANOVA, 并进行 FDR 校正。

(3) 组间差异显著性检验—两组比较: 基于 OTU 丰度表, 使用 R 软件进行 Wilcoxon 秩和检验、Wilcoxon 符号秩检验、Student's t 检验或 Welch's t 检验, 并进行 FDR 校正。

## 2.8 环境因子关联分析

(1) Mantel\_Test 分析: 基于 OTU 丰度表以及环境因子数据, 使用 R 软件 `vegan` 的软件包进行 Mantel test 分析, 并根据所求  $r$  值和显著性水平  $p$  值来判断环境因子和微生物群落分布的相关性。

(2) CCA/RDA 分析: 基于 OTU 丰度表, 以及环境因子数据, 首先使用 R 软件进



行去趋势对应分析，即 DCA(Detrended correspondence analysis)分析，根据梯度值确定线性模型(RDA)和单峰模型(CCA)哪个最合适（DCA 分析结果中 Axis length 的前 4 个轴中最大的值如果大于 4.0，应该选 CCA，如果 3.0-4.0 之间，选 RDA 和 CCA 均可，如果小于 3.0，RDA 的结果要好于 CCA），然后使用进行 CCA 或 RDA 分析并绘图。

(3) 相关性热图分析：基于 OTU 丰度表以及环境因子数据，使用 R 软件进行群落结构与各个环境因子之间的关联分析并绘制热图展示。

(4) VPA 分析：利用 R 软件分析各环境因子对微生物群落分布的解释量，可得到造成微生物群落分布差异的各环境因素排除其他因素影响后对群落物种分布构成差异的解释度。

(5) 排序回归分析：基于 OTU 丰度表以及环境因子数据，进行环境因子排序回归分析，根据 PCA/PCoA/NMDS 分析结果，以各样本在某一轴（如 PC1 轴）上的分值为 x 轴，以该样品对应的环境因子（如 pH、温度等）为 y 轴做散点图，并进行线性回归（Linear Regression）。

## 2.9 网络与预测分析

(1) 共线性网络分析：基于 OTU 丰度表，使用 R 软件，进行对不同样本间的物种丰度信息进行共线性网络分析。

(2) 共线性网络分析：基于 OTU 丰度表，使用 R 软件，使用 R 软件计算 pearson 或 spearman 相关系数，获得样本内或样本组内的物种之间的相互关系，然后用展示物种间相互关系的可视化软件进行物种相互作用 network 构建及展示。

(3) Random\_Forest 分析：基于 OTU 丰度表，使用 R 软件的 randomForest 软件包，进行建立预测模型、关键物种及自定义验证集取得预测结果分析。

## 2.10 16S 功能预测分析

(1) COG 功能分析: 通过 PICRUST(<https://picrust.github.io/picrust/>)对 OTU 丰度表进行标准化, 去除 16S marker gene 在物种基因组中的拷贝数的影响, 然后通过每个 OTU 对应的 greengene id, 比对到 COG 数据库, 获得 OTU 对应的 COG 家族信息, 计算各 COG 的丰度; 根据比对到的 COG 库的 COG 编号可以从 eggNOG 数据库中解析到各个 COG 的描述信息, 及其功能信息, 从而得到功能丰度谱, 进而进行后续分析。

(2) KEGG 功能分析: 通过 PICRUST 对 OTU 丰度表进行标准化, 去除 16S marker gene 在物种基因组中的拷贝数的影响, 然后通过每个 OTU 对应的 greengene id, 比对到 KEGG 数据库, 根据比对到 KEGG 数据库的信息, 可以获得 KO、Pathway、EC 信息, 并根据 OTU 丰度计算各功能类别的丰度, 分别得到各样本在不同分类水平的丰度表, 进而进行后续分析。

## 3. 试剂、耗材及仪器

### 3.1 试剂列表

名称	厂家	货号	备注
MOBIO PowerSoil® DNA Isolation Kit（视样本类型而定）	MOBIO	12888-100	MOBIO Laboratories, Carlsbad, CA, USA
Premix Taq（EX Taq Version 2.0 plus dye）	TaKaRa	RR902A	Takara Biotechnology, Dalian Co. Ltd., China
Primer	Invitrogen		Carlsbad, CA, USA
琼脂糖	浩玛生物	111860	
Quant-iT™ Broad-Range DNA Assay Kit	Invitrogen	Q32853	Invitrogen, Carlsbad, CA, USA
DL15000 DNA Marker	TaKaRa	3582A	
100bp DNA ladder	TaKaRa	3422A	
DL2000 DNA Marker	TaKaRa	3247A	
E.Z.N.A.® Gel Extraction Kit	Omega	D2500-00	Omega, USA
NEBNext® Ultra™ DNA Library Prep Kit for Illumina®	New England Biolabs	E7370L	New England Biolabs, MA, USA

### 3.2 常用耗材

名称	厂家	货号	备注
0.2mlPCR(薄壁)管(平盖)	Axygen	PCR-02-C	
0.5mlPCR 透明薄壁管（平盖）	Axygen	PCR-05-C	
1.5mL 离心管	Axygen	MCT-150-C	
2mL 离心管	Axygen	MCT-200-C	

### 3.3 常用仪器

名称	厂家	货号	备注
Dack reader	Clare Research	Chemical DR-46B	
Qubit™ fluorometer	Invitrogen	Qubit® Fluorometer	2.0 Invitrogen , Carlsbad , CA, USA
微型离心机	Baygene	Baygene BG-Qspin#8482	
涡旋混合器	其林贝尔	QL-901	
NanoDrop™ One	Thermo Scientific	Fisher ND-ONE-W	Thermo Fisher Scientific , MA, USA
电泳仪	北京六一	DYY-6C	
凝胶成像仪	Tanon	Tanon 4100	
PCR 仪	BioRad	S1000	Bio-Rad Laboratory, CA, USA
冷冻离心机	Eppendorf	5417R	